Abdominal CTA Image Analisys Through Active Learning and Decision Random Forests: Aplication to AAA Segmentation.

Josu Maiora Manuel Graña

Abstract—Abdominal Aortic Aneurysm (AAA) is a local dilation of the Aorta that occurs between the renal and iliac arteries. The weakening of the aortic wall leads to its deformation and the generation of a thrombus. Recently developed treatment involves the insertion of an endovascular prosthetic (EVAR), which has the advantage of being a minimally invasive procedure but also requires monitoring to analyze postoperative patient outcomes using 3D Contrast Computerized Tomography Angiography (CTA) imaging procedures. In order to effectively assess the changes experienced after surgery, it is necessary to segment the aneurysm in the CT volume, which is a very time-consuming task. Here we provide results of a novel active learning approach for the semi-automatic detection and segmentation of the lumen and the thrombus of the AAA, which uses image intensity features and discriminative Random Forest classifiers.

Index Terms—edical Image, Segmentation, Active Learningedical Image, Segmentation, Active LearningM

I. INTRODUCTION

Abdominal Aortic Aneurysm (AAA) is a local dilation of the Aorta that occurs between the renal and iliac arteries. The weakening of the aortic wall leads to its deformation and the generation of a thrombus. Generally, an AAA is diagnosed when the minimum anterioposterior diameter of the aorta reaches 3.0 cm [1]. The majority of aortic aneurysms are asymptomatic and without complications. Aneurysms that cause symptoms have a higher risk of rupture. Abdominal pain or back pain are the two main clinical features suggestive of either the recent expansion or leakage. The complications are often life threatening and can occur in a short space of time. Therefore, the challenge is to diagnose before the onset of symptoms. Asymptomatic aneurysms are often detected incidentally[2].

The prevalence of AAA depends on various risk factors, as advancing age, family history, male gender, and tobacco use. According to the ACC/AHA guidelines (Hirsch et al., 2006) the prevalence of AAA 2.9 to 4.9 cm in diameter varies with age and sex, ranging from 1.3 % up to 12.5 % for men aged in the [45,54] and [75,84] year intervals, respectively. Prevalence figures for women in the same age intervals are 0% and 5.2 %, respectively. In these studies the AAA was defined as an aortic diameter ≥3cm. If clinically important aneurysms are only taken into account (AAA measuring ≥ 4 cm in diameter) the indicated prevalence would be lower.

Josu Maiora is with the Electronics and Telecommunication Department, University of the Basque Country.
Manuel Graña is with the Computational Intelligence Group, University of the Basque Country.

The statements about the natural course of the disease and the risk of rupture are mainly based on the results of the randomized observational studies that compared immediate repair with surveillance for small AAA[3]. The evidence report prepared for the Agency for Healthcare Research and Quality (AHRQ) analyzed these and further observational studies [4]. They came to the conclusion that the annual risk of rupture is 1% or lower for AAA less than 5.5 cm. The 1-year risk of rupture increases with aneurysm size and may exceed 10% in individuals with AAA > 6 cm. For AAA that attain a size of > 8 cm, the risk may exceed 25 % at 6 months. Female sex, higher mean arterial blood pressure, current smoking, and poor lung function increase the risk of aneurysm rupture in addition to the size of initial AAA diameter.

3D Contrast Computerized Tomography Angiography (CTA) is the preferred imaging method because it allows minimally invasive visualization of the Aorta’s lumen, thrombus and calcifications. Even though several segmentation methods for vascular structures have been developed [5] [6], the segmentation of the AAA thrombus is still a challenging task due to the low contrast of signal intensity values between the aneurysm thrombus and its surrounding tissue as can be appreciated in Fig. 1. Several AAA thrombus segmentation methods have been recently developed. The method by De Brujine et al.[7] is an interactive contour tracking method for axial slices; Olabarriaga et al. [8] employ a deformable model approach based on a nonparametric statistical grey-level appearance model to determine the deformable model adaptation direction starting from a lumen contour shape interactive segmentation; Zhuge et al. [9] present a level-set segmentation based on a parametric statistical model; Demirci et al. [10] propose a deformable B-spline parametric model based on a nonparametric intensity distribution model and; Freiman et al. [11] apply an iterative model-constrained graph-cut algorithm.

These methods all involve a significant user interaction or initialization, and there are small differences between them. They define a dynamics (or an optimization) of auxiliary variables associated with the pixels. Each variable is updated depending on a linear combination of variables from neighboring pixels, as well as some kind of nonlinear operation.

In this paper we propose a machine learning based method. Machine Learning has yielded superior performance in the Berkeley Segmentation Benchmark [12] comparing to other
methods on the segmentation of natural images, using a large variety of image features. In [8] the k-NN classifier is applied to learn the energy magnitude of the deformable model external forces on intensity profiles drawn along a thrombus boundary normal line. Collecting the intensity profiles for training is a delicate issue, because they must be centered on the thrombus boundary. The method needs, therefore, an accurate expert delineation of such boundary. The force magnitude is proportional to the likelihood of the intensity profile according to the k-NN probabilistic model, which is heavily dependent on the actual data used to build the system.

In order to obtain the most efficient training with the smallest possible training set, we apply an active learning approach. The main philosophical issue behind the approach is that we can not ensure that our training data sample will be representative of the actual data conditions. Therefore, we assume that the image segmentation will need to be retrained often. The aim is to show that this retraining can be performed efficiently in time and superior performance. We perform the classification of the pixels of the entire image with random forest (RF) classifier. Several authors have been developing RF based image segmentation techniques in the last years. Lempitsky et al. [13] have used the binary RF to automatically delineate the myocardium in 3D ultrasound (US) of adult hearts. Yi et al. [14]segmented the three main brain tissues from MRI volumes. Geremia et al. [15] segmented multiple sclerosis in multi-channel brain MR images. Yaqub et al. [16] proposed a weighted RF technique in which weights are assigned to trees during testing depending on their strength to classify a new test case. On the other hand, Criminisi et al. [17], [18] have used the standard RF technique to automatically detect several organs in CT volumes, finding a 3D bounding box around each organ.

II. METHODS

This section describes the active learning procedure and the random classification forest we use for the segmentation process of the AAA.

Our detection and segmentation problem is a multiclass classification of voxel samples into aortic lumen, thrombus, bones (column) and background. We perform the classification with a supervised method: RF. We build the training data in an iterative active learning process (Fig. 2).

A. Active learning

In the current state of the art, the use of statistical learning models is a common practice for other research areas like remote sensing; Support Vector Machines (SVM) [19] or neural networks[20] algorithms are widely used for the classification. However, the performances of supervised algorithms strongly depend on the information gain provided by the data used to train the classifier. This makes the construction of the training set a cumbersome task requiring extensive manual analysis of the image. This is typically done by visual inspection of the scene and successive labeling of each sample. Consequently, the training set is highly redundant and training phase of the model is significantly slowed down. Besides, noisy pixels may
interfere the class statistics, which may lead to poor classification performances and/or over-fitting. For these reasons, a training set should also be kept as small as possible and focused on those pixels effectively improving the performance of the model. Therefore a desirable training set must be constructed in a smart way, meaning it must represent correctly the class boundaries by sampling discriminative pixels. In the machine learning literature this approach to sampling is known as active learning[16].

Active learning focuses on the interaction between the user and the classifier. The model returns to the user the pixels whose classification outcome is the most uncertain. After accurate labeling by the user, pixels are included into the training set in order to reinforce the model[15]. This way, the model is optimized on well-chosen difficult examples, maximizing its generalization capabilities.

B. Random Forest Classifiers

We may summarize the classification task as: Given a labeled training set learn a general mapping which associates previously unseen test data with their correct classes.

The random forests (RF) machine learning algorithm is a classifier [21] that encompasses bagging [22] and random decision forests [23][24] and is used in a variety of applications [25]. RF became popular due to its simplicity of training and tuning while offering a similar performance to boosting. It is a large collection of decorrelated decision trees, which are ideal candidates to capture complex interaction structures in data. RF is supposed to be resistant to over-fitting of data if individual trees are sufficiently deep. Consider a RF collection of tree predictors

\[ h(x; \psi_u), u = 1, ..., U, \]

where \( x \) is a random sample of \( d \)-dimensions associated to random vector \( X \) and \( \psi_u \) independent identically distributed random vectors. Given a dataset of \( N \) samples, the bootstrap training sample corresponding to tree \( h(x; \psi_u) \) is used to grow it by recursively selecting a subset of random dimensions \( d \) such that \( d \ll d \) and picking the best split of each node based on these variables. Unlike conventional decision trees, pruning is not required. Let us denote:

\[ \hat{c} = \text{majority vote}\{C_u(x)^n\} \] (1)

To make a prediction for a new sample \( x \), the trained RF could then be used for classification by majority vote among the trees of the RF as shown in Eq. (1), where \( C_u(x) \) is the class prediction of the \( u \)-th RF tree. The important parameters of the RF classifier were set as follows in this case. The number of trees in the forest should be sufficiently large to ensure that each input class receives a number of predictions: we set it to 200. The number of variables randomly sampled at each branch: we set it to 5.

C. Feature set construction through active learning

Feature Selection: We build a minimal set of manually labeled voxels from the CT images. A feature set is a group of features of the same type but with different dimensions and extracted from image locations given by a window around a pixel of interest. That is, we consider spatial information around the pixel besides its own intensity information. This information comes from the result of linear and/or non-linear filtering performed on this window. For instance, the Sobel filter features correspond to extracting a window over the gradient image computed applying the Sobel convolution kernel. The maximum corresponds to computing the maximum value inside the window. Initially we choose a wide variety of the most common image intensity features (Table I) as well as different radius values (1,2,4...2^\text{nd}) around the voxel of interest.

However the training set build this way has an huge number of features, so that it is convenient to get a less expensive computational task. Since many of the possible features provide relatively poor information gain, the probability to select a good feature is low. To get a more efficient algorithm we compute the information gain provided by each feature and radius value. We evaluate the worth of a feature by measuring the information gain with respect to the class,

\[ IG(Class, Feature) = H(Class) - H(Class|Feature) \]

where \( H \) is the entropy of the corresponding distribution. \( Class \) is the \textit{a priori} distribution of the class, the \textit{Class} conditioned to the \textit{Feature} corresponds to the \textit{a posteriori} distribution resulting from the classifier.

Then we discard features that provide less information gain than 0.5. In Table II we show the finally selected Image features to build the training set.

Then we choose the features that give more information, we get a smaller set that produces similar classification results.

Adding labeled data and active learning: As observed in Fig 1, the contour of the aneurysm thrombus shows higher classification uncertainty (pixel intensity similar to surrounding tissue). Thus, we might decide to collect and label additional data precisely in those low-confidence regions [26]. This has the effect of refining the classification posterior and increasing its accuracy. As expected, a guided addition of further labeled data in regions of high uncertainty increases the overall predictor confidence.

Handling multiple classes: Although our final goal is the segmentation of the thrombus, it is desirable to get the segmentation of other anatomical structures or organs simultaneously. semi-supervised forests are tree-based models that

<table>
<thead>
<tr>
<th>Image Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian smoothing</td>
</tr>
<tr>
<td>Sobel filter</td>
</tr>
<tr>
<td>Hessian</td>
</tr>
<tr>
<td>Difference of gaussians</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>Laplacian</td>
</tr>
</tbody>
</table>
can handle easily multiple (> 2) classes. Following the same procedure we perform a four-class (lumen, thrombus, bones, background) segmentation experiment. Out of hundreds of points only 12 are labeled in each iteration with their respective classes. Conventional one-vs-all SVM and other neural network classifications results in harder class assignments as we will demonstrate later.

D. Morphological Operators

Morphological operators [27] are nonlinear filters, convolving masks over the image applying lattice operators that can be used for edge detection, smoothing, removing noise and even detect shapes. Dilation, erosion, closing, and opening are the basic morphological operators. They need an structuring element, which is a smaller image that can be associated with a shape. We are concerned with binary images, whose pixels have values 0/1. The structuring element can be of any size and has an arbitrary structure that can be represented by binary values. We will use a circular region of \( d \) diameter.

The binary erosion of \( A \) by \( B \), denoted \( A \odot B \), is defined as the set operation \( A \odot B = \{ z \mid (B)_z \subseteq A \} \). In other words, it is the set of pixel locations \( z \), where the structuring element translated to location \( z \) overlaps only with foreground pixels in \( A \).

The binary dilation of \( A \) by \( B \), denoted \( A \oplus B \), is defined as the set operation: \( A \oplus B = \{ z \mid (B)^c_z \cap A \neq \emptyset \} \) where \( (B)^c \) is the reflection of the structuring element \( B \). In other words, it is the set of pixel locations \( z \), where the reflected structuring element overlaps with foreground pixels in \( A \) when translated to \( z \).

Dilation and erosion are often used in combination to detect sub-images or image components. The definition of a morphological opening of an image is an erosion followed by a dilation, using the same structuring element for both operations: \( X_B = (X \odot B) \oplus B \). The morphological closing of an image is the reverse: it consists of a dilation followed by an erosion with the same structuring element: \( X^B = (X \oplus B) \odot B \).

III. EXPERIMENTAL SETUP

Datasets. 8 datasets were used to experimentally test the proposed technique. Each dataset consists in real human contrast-enhanced datasets of the abdominal area obtained from a LightSpeed16 CT scanner (GE Medical Systems, Fairfield, CT, USA) with 512x512 pixel resolution on each slice. Each dataset consists of between 312 and 560 slices and 0.725x0.725x0.8 mm spatial resolution corresponding to patients who suffered Abdominal Aortic Aneurysm.

Validation. We have designed an experiment to test our method with a human operator. The operator is asked first to select a random slice of the central part of the thrombus. Second, the operator performs 8 active learning iterations. At each active learning iteration the operator takes 12 point samples, 2 corresponding to lumen and bones, and 4 corresponding to background and thrombus. Each active learning iteration adds previously misclassified data samples to the training set, and the classifier is retrained on the incremented training dataset. The performance is recorded using Dice’s coefficient (DC)

\[
D(A, B) = \frac{|A \cap B|}{|A| + |B|},
\]

computed over all the volume. The DC is, therefore, a measure of the generalization of the classifier learned on one slice to the remaining slices in the volume. We use a manual segmentation of each slice as ground truth. We compute the DC after each iteration.

Implementation. In our experiments we build the feature set with Advanced Weka Segmentation, which is a plugin of Fiji medical image processing software, running over the Weka data mining software [28].

We train four different classifiers over the set of features: Support Vector Machines (SVM) with linear kernel and Radial Basis Function (RBF) kernel, Multi Layer Perceptron (MLP) and Random Forest. In this step, we use standard parameter settings for the classifiers. Three-fold cross validation is used in every experiment. We perform 10 repetitions in order to avoid testing errors. As we get the best results for Random Forests, next we proceed to test the sensitivity of the forest configurations. Each dataset consists in real human contrast-enhanced datasets of the abdominal area obtained from a LightSpeed16 CT scanner (GE Medical Systems, Fairfield, CT, USA) with 512x512 pixel resolution on each slice. Each dataset consists of between 312 and 560 slices and 0.725x0.725x0.8 mm spatial resolution corresponding to patients who suffered Abdominal Aortic Aneurysm.

IV. RESULTS

Results presented in this section aim at evaluating the feature extraction process, voxel classification after the consecutive iterations and finally the segmentation results.
Table III
CROSS-VALIDATION RESULTS OVER THE ABDOMINAL IMAGE FEATURES COMPUTED FROM THE CT DATASETS FOR ABDOMINAL IMAGES SEGMENTATION

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear SVM</td>
<td>0.72</td>
<td>0.75</td>
<td>0.97</td>
</tr>
<tr>
<td>RBF SVM</td>
<td>0.77</td>
<td>0.80</td>
<td>0.98</td>
</tr>
<tr>
<td>BP-MLP</td>
<td>0.73</td>
<td>0.71</td>
<td>0.97</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.91</td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Figure 3. Accuracy of the segmentation as a function of the forest parameters: number of trees and depth.

In Table III we show the results for accuracy, sensitivity and specificity for the four different classifiers on the selected slice after the active learning iterations. The final result is the average outcome of the 10 repetitions on the test data. We obtain the best results for RF followed by SVM with RBF kernel, while MLP and SVM with linear kernel give us the less accurate results.

In Fig. 3 segmentation results are evaluated for each combination of RF parameters to compute the accuracy surface response to variations in RF parameters. The figure shows that for a fixed depth, increasing the number of trees and the depth leads to a more accurate classification. The increase in performance stabilizes around number of trees = 80 and depth = 20.

Abdominal CT slices segmentation results after each of the first five active learning iterations and the last one are showed in 4. It can be appreciated how the thrombus detection increases decreasing the false detections on the surrounding tissues with confounding intensity values. Segmentation improves significantly and in the 8th iteration, the bones (backbone and rib), lumen and thrombus are clearly distinguished.

The classifier built on the training set corresponding to the image features of one slice has a good performance detecting and segmenting the anatomical structures in several consecutive slices (Fig. 5). We can observe that the segmentation quality has minimum variations from one slice to the next.

Figure 4. Segmentation results after each of the first 5 and the last active learning iterations of the same slice and patient after RF classification process. Lumen (darkest circle in the center), thrombus (circle around lumen), and bones (backbone and rib) are distinguished. (a-e) show the first 5 interactions, while (f) shows the segmentation after the final iteration.

Figure 5. 6 consecutive abdominal segmented images of the same patient after RF classification process. Lumen (darkest circle in the center), thrombus (circle around lumen), and bones (backbone and rib) are distinguished.

Thrombus segmentation is preserved, and the false positives are distributed in similar way.

The most computationally expensive operation is the feature stack extraction for the entire volume, which took about 14 min of CPU time to complete (Intel i7 processor, 8MB of RAM). Once we have the feature stack, the classification process took about 8 minutes.

The final step of the system applies a morphological closing operator to get rid of the false positives, obtaining the mask of the aneurysm thrombus. Fig. 6 shows the results of the closing on several consecutive slices of the same patient after classification with the RF trained in 8 active learning iterations. We observe little variation between consecutive slices.

The evolution of the DC values after each active learning
Figure 6. 6 consecutive abdominal aortic aneurysm thrombus segmented images of the same patient after RF classification process and morphological opening operation.

Figure 7. Dice similarity measures after each iteration.

iteration is shown in the figure 7. After the first two iteration DC values are still relatively low, but after the third iteration there is a steep increase above 0.8. We observe that the segmentation quality reaches a maximum after 6 active learning iterations.

A 3D volume rendering for both aortic lumen and thrombus is shown in Figure 8. We can observe that the thrombus segmentation is more homogeneous in the middle than in the upper and lower parts. The reason is that we have used a slice from the central part to train the classifier, consequently its performance is better in the nearby slices than in slices located at the volume extremes. Because the contrast agents produce a strong image intensity contrast, the segmentation of the lumen is homogeneous all over the slices.

V. CONCLUSION AND FUTURE WORKS

In this paper we propose an Active Learning approach for the segmentation of the thrombus in EVAR followup images. First, we use these Active Learning techniques to select the optimal feature sets, evaluating the information gain of a variety of intensity based features, to train the classifier and perform the voxel-based segmentation. Active Learning allows to obtain near optimal segmentation results with minimal intervention of the human operator in a very fast semi-automated process.

We compare the Random Forest (RF) classifier with other state-of-the-art classifiers and neural networks. Results show that RF achieve more accurate results and thus provide a efficient tool for discriminating voxels corresponding to specific anatomical structures in abdominal CT images. Specifically, we discriminate efficiently the voxels corresponding to the thrombus.

Our method is being currently tested on real human datasets and results are promising. Accurate segmentation is obtained in areas where it is difficult to distinguish the thrombus from surrounding structures and is a good input for a generative model that would improve the segmentation quality.

Future work will be oriented to improve the training set, fine-tune the parameters of the process for a large number of datasets and validate the segmentation by comparison with manual segmentation and other methods.

REFERENCES


